LHCOPN and LHCONE update

ATCF6 Krabi - 22 November 2022
edoardo.martelli@cern.ch

Agenda

- CERN Tier-0

- LHCOPN

- LHCONE

- Network R&D

CERN Tier-0 update

Highlights

LHC
- Run3 in full swing
- Run4 postponed to 2029

Concerns on electricity supplies
- preparing for unexpected power cuts
    during the winter
- extending maintenance periods to
    decrease running costs

Computing
- storage and computing reliable and growing
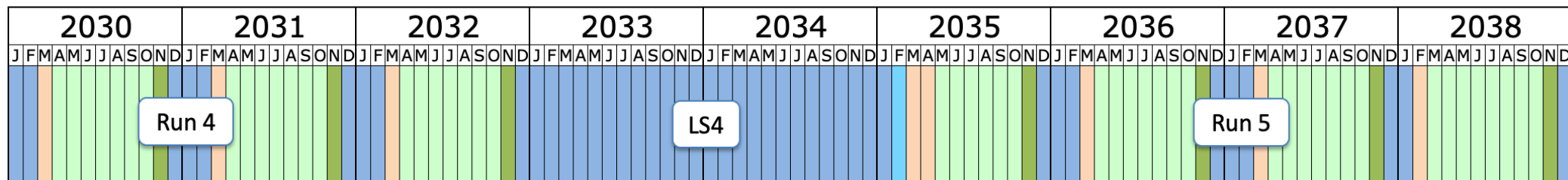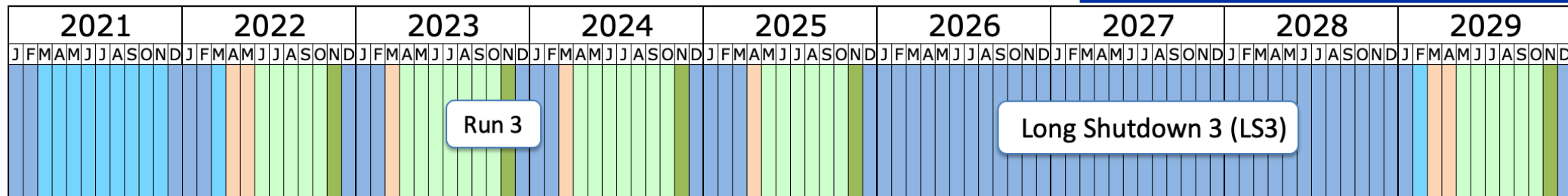- new datacentre being built

Network
- growing in use and capacity
- accelerating adoption of 400Gbps

Construction of CERN Science Gateway

LHC schedule

Measures to mitigate impact of electricity costs

❑ **2022 YETS** (Year-End-Technical-Stop) **starts 2 weeks earlier** (28 Nov)
→ mark of social responsibility; savings

❑ In **2023, the accelerator complex operation will be reduced by 20%** (i.e. YETS will be extended from 15 to 19 weeks)
→ mark of social responsibility; savings

❑ **Package of measures for 2024-2032 being prepared** (preliminary discussions at the Sept Council → will continue in Dec):
efforts from all stakeholders: savings from CERN's activities; savings on personnel costs; possible indexation of Member and Associate Member States contributions beyond 2% "ceiling".
Aim is to bring cumulative budget deficit to ~ zero in the early 2030s, so as to be able to start investment in a new big facility at CERN

❑ Reduced power configurations prepared to cope with possible EDF **load shedding**

❑ **Crisis procedures to face** total (prolonged) **blackouts** being prepared

❑ Measures to **reduce energy consumption on CERN campus** being implemented

PCC (Prevessin Computer Centre) status
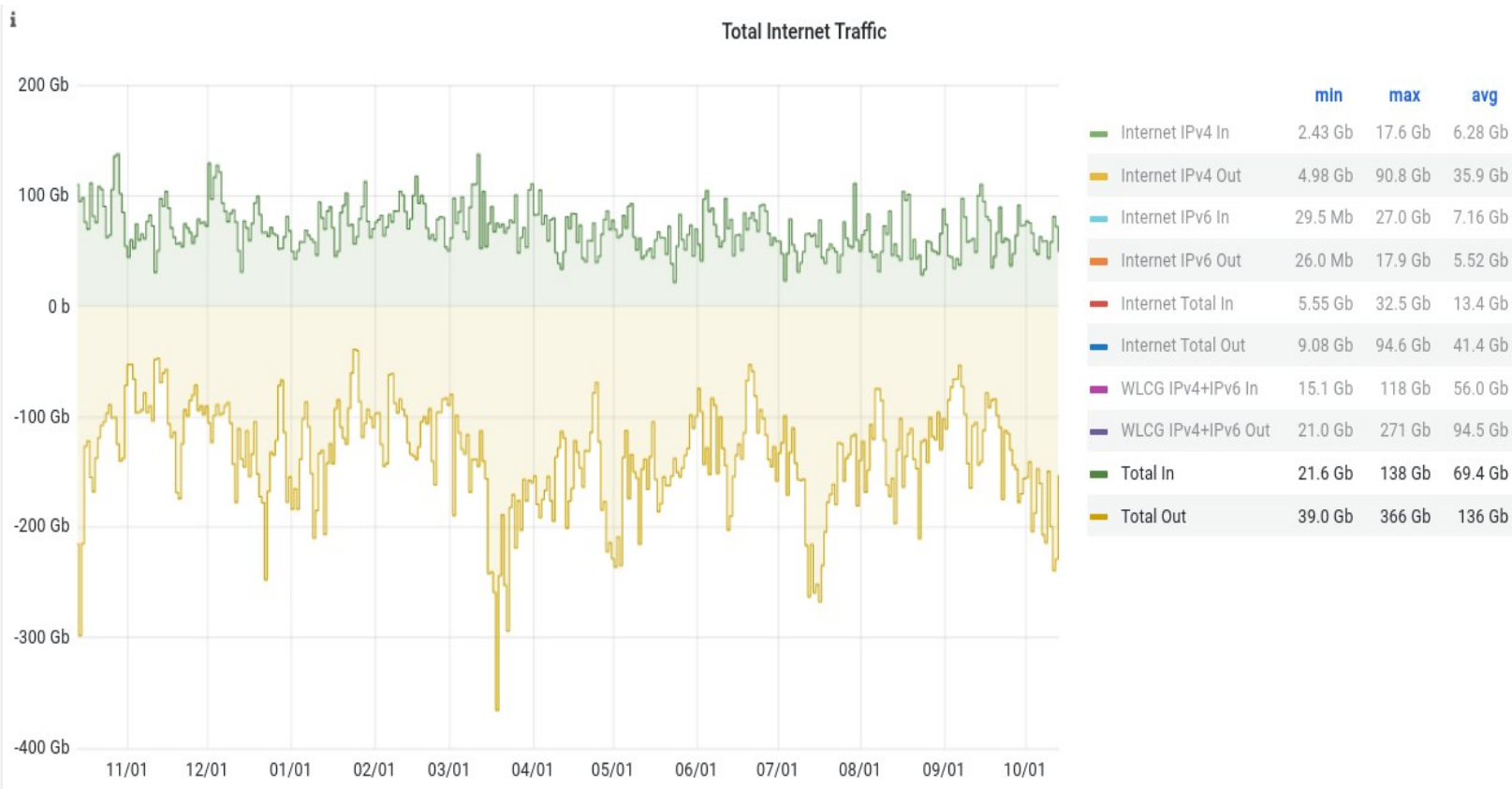
2022

- End 2022: Civil engineering and structural work to be completed
- December: Tender for first installation of servers to be adjudicated

2023

- Summer: Delivery and installation in PCC planned
- 3rd quarter: Data Centre ready for commissioning
- End 2023: Inauguration Ceremony foreseen

CERN External Network Traffic

Total Internet Traffic

	min	max	avg
Internet IPv4 In	2.43 Gb	17.6 Gb	6.28 Gb
Internet IPv4 Out	4.98 Gb	90.8 Gb	35.9 Gb
Internet IPv6 In	29.5 Mb	27.0 Gb	7.16 Gb
Internet IPv6 Out	26.0 Mb	17.9 Gb	5.52 Gb
Internet Total In	5.55 Gb	32.5 Gb	13.4 Gb
Internet Total Out	9.08 Gb	94.6 Gb	41.4 Gb
WLCG IPv4+IPv6 In	15.1 Gb	118 Gb	56.0 Gb
WLCG IPv4+IPv6 Out	21.0 Gb	271 Gb	94.5 Gb
Total In	21.6 Gb	138 Gb	69.4 Gb
Total Out	39.0 Gb	366 Gb	136 Gb

Numbers:

Sent out ~536 PB in the last 12 months
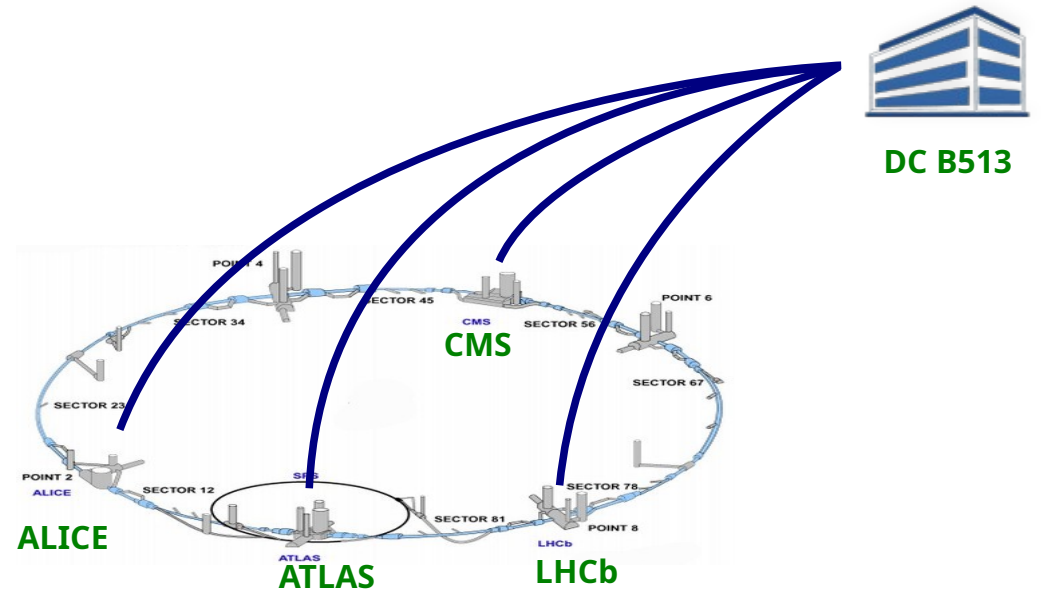
+34% compared to previous year (398PB)

Ref: https://monit-grafana.cern.ch/d/cScW82Tnz/00-overview?orgId=14&var-source=long_term&var-bin=1d&from=now-1y&to=now
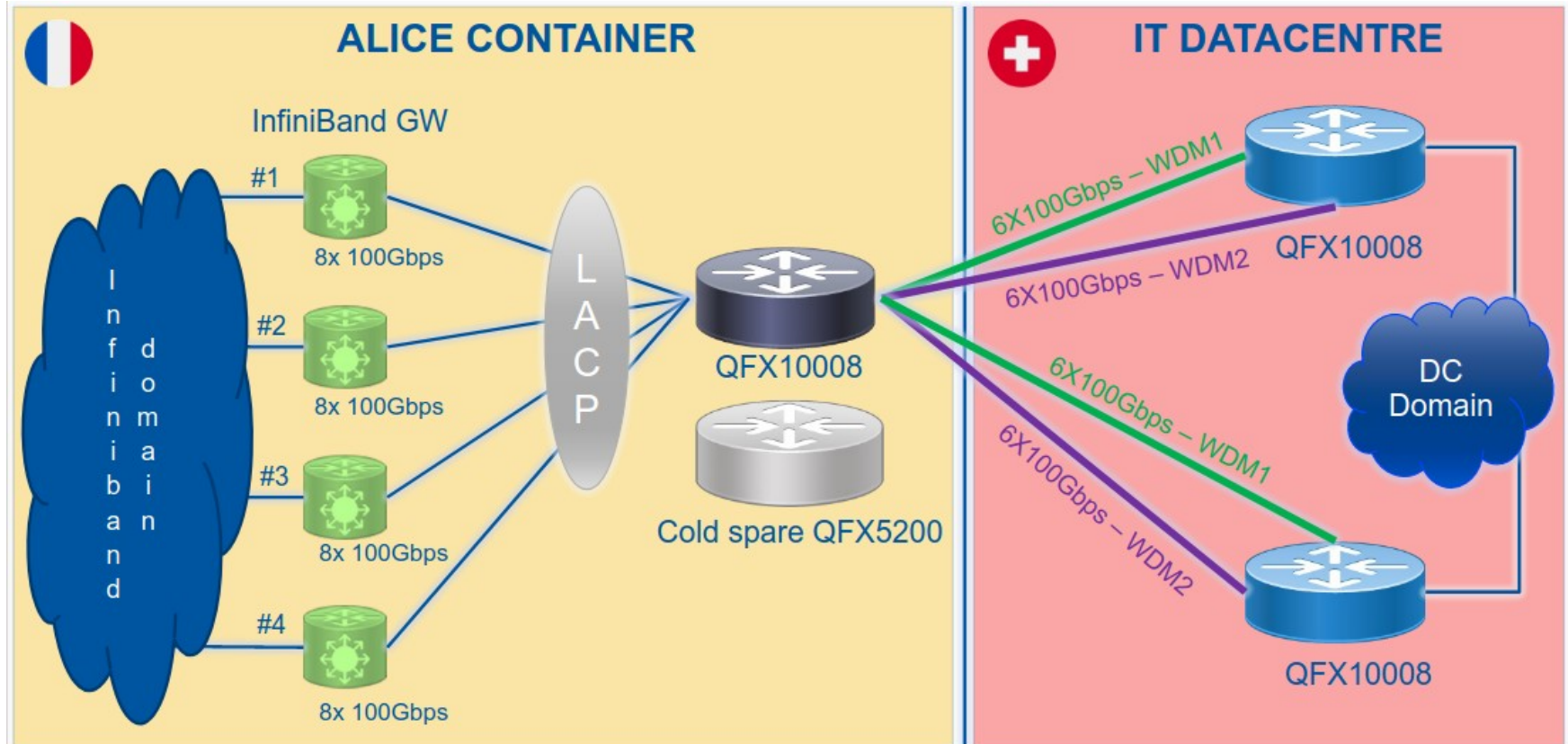
Experiments' DAQ lines to IT data-centre

Capacity in place for Run3:
- ALICE: 2.4 Tbps
- LHCb: 400 Gbps
- CMS: 400Gbps
- ATLAS: 200Gbps
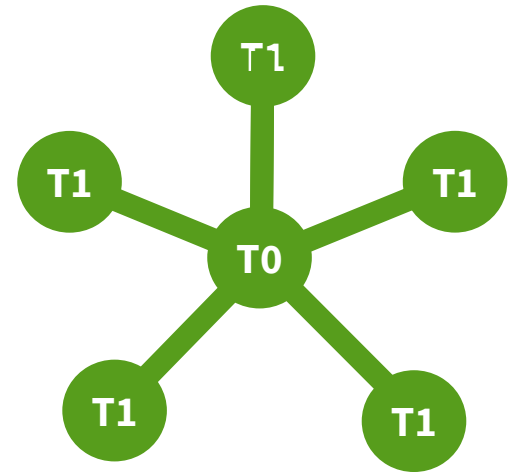
ALICE DAQ: 2.4Tbps on two fibres with DWDM

LHCOPN

LHCOPN

Private network connecting Tier0 and Tier1s

Secure:
- Dedicated to LHC data transfers
- Only declared IP prefixes can exchange traffic
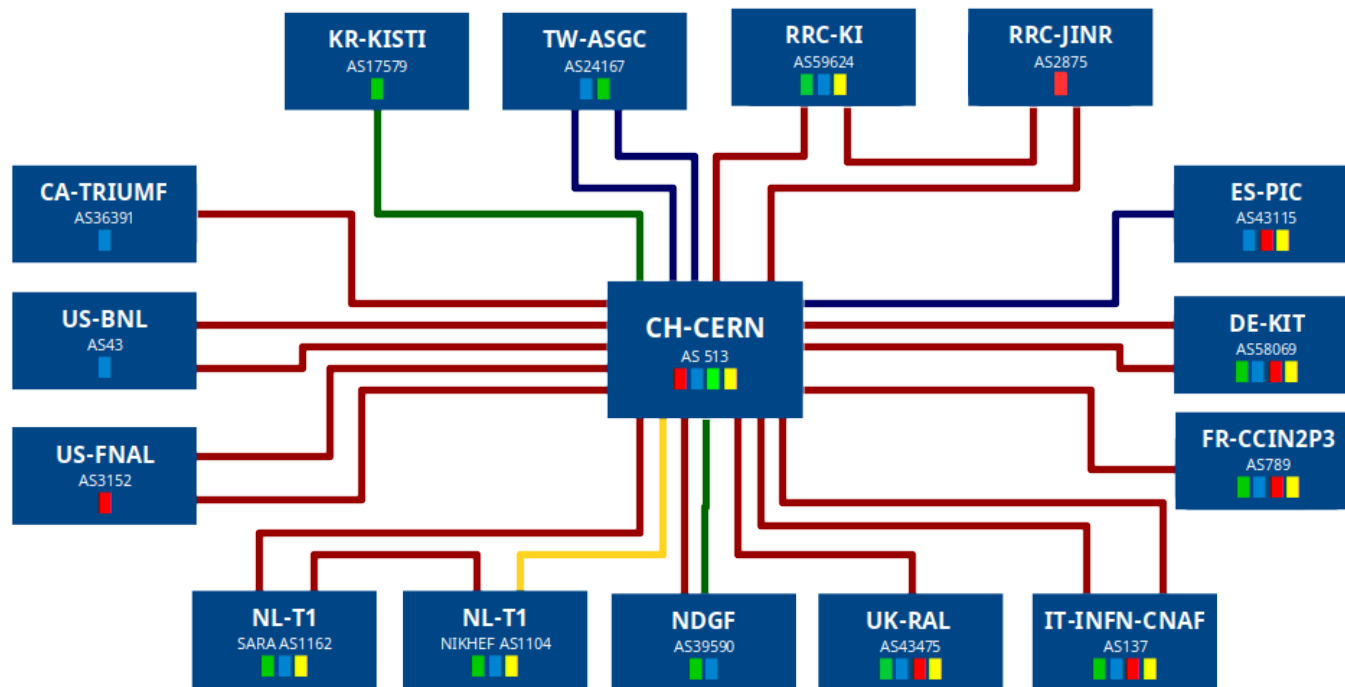- Can connect directly to Science-DMZ,
  bypass perimeter firewalls

Advanced routing:
- BGP communities for traffic engineering
- load balancing

Numbers

- 14 Tier1s + 1 Tier0
- 12 countries in 3 continents
- Dual stack IPv4-IPv6
- 1.9 Tbps to the Tier0

https://twiki.cern.ch/twiki/bin/view/LHCOPN/OverallNetworkMaps

Latest news

KR-KISTI
- Upgrading connectivity to Amsterdam to 2x100Gbps
- Tier1 link to CERN will be upgraded to 100Gbps

NLT1:
- New 400G connection to NIKHEF. SURFsara still using the 100G link
- SURF will provide a redundant path for the two T1s

IT-INFN-CNAF:
- Following the recent upgrade of the GARR network, CNAF has activated traffic load-balancing
    over their 2x 100G LHCOPN links to CERN

ES-PIC:
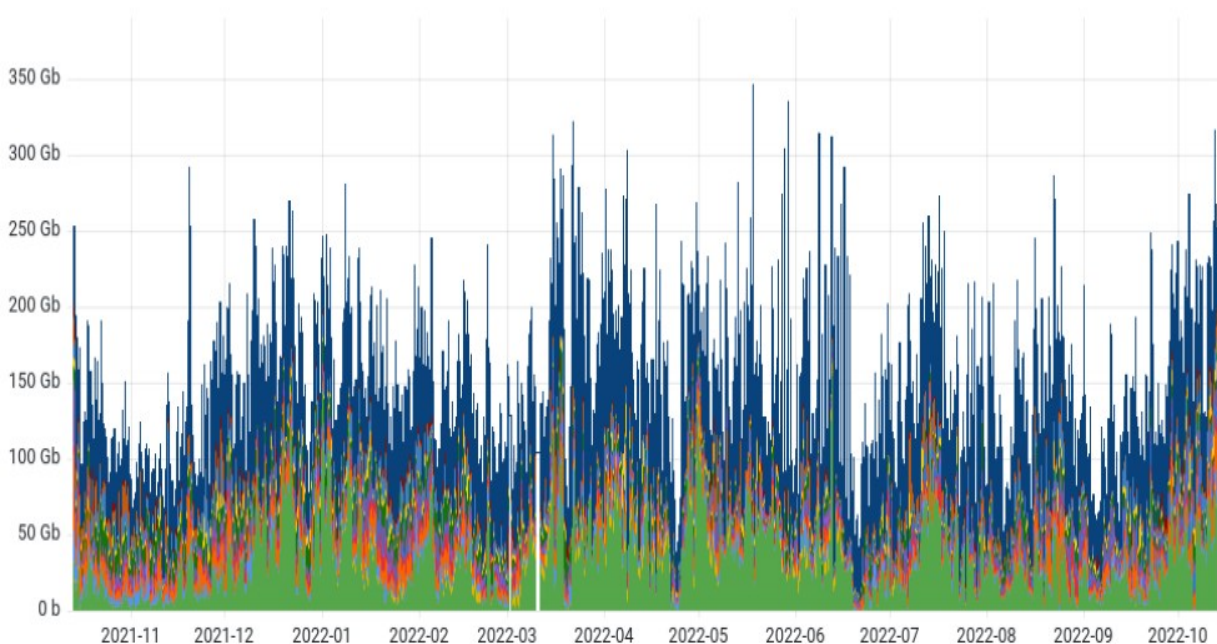- Upgrade to 100Gbps ready.to go in production

Changes:
- ASGC Taiwan has resigned the Tier1 role
- New Tier1 application from IHEPCC China and Swierk Poland

LHCOPN

LHCOPN Traffic – last 12 months

LHCOPN Total Traffic (CERN -> T1s)

	Mean	Max
Outgoing DE-KIT	50.8 Gb	218 Gb
Outgoing KR-KISTI	1.52 Gb	12.7 Gb
Outgoing RU-T1	4.75 Gb	46.0 Gb
Outgoing FR-IN2P3	11.0 Gb	73.2 Gb
Outgoing NDGF	4.53 Gb	37.3 Gb
Outgoing NL-T1	3.10 Gb	70.3 Gb
Outgoing TW-ASGC	1.71 Gb	9.65 Gb
Outgoing IT-INFN-CNAF	7.91 Gb	98.4 Gb
Outgoing UK-RAL	8.15 Gb	35.3 Gb
Outgoing CA-TRIUMF	4.36 Gb	52.0 Gb
Outgoing US-BNL	10.0 Gb	81.2 Gb
Outgoing US-FNAL	4.95 Gb	82.3 Gb
Outgoing ES-PIC	2.84 Gb	11.8 Gb
Total	116 Gb	346 Gb

Numbers:

Moved ~457 PB in the last 12 months

+34% compared to previous year (341PB)

Peak at ~391Gbps

Ref:https://monit-grafana.cern.ch/d/HreVOyc7z/all-lhcopn-traffic?orgId=14

LHCOPN growth

Gbps - Y-Axis label

Run1   Run2   Run3
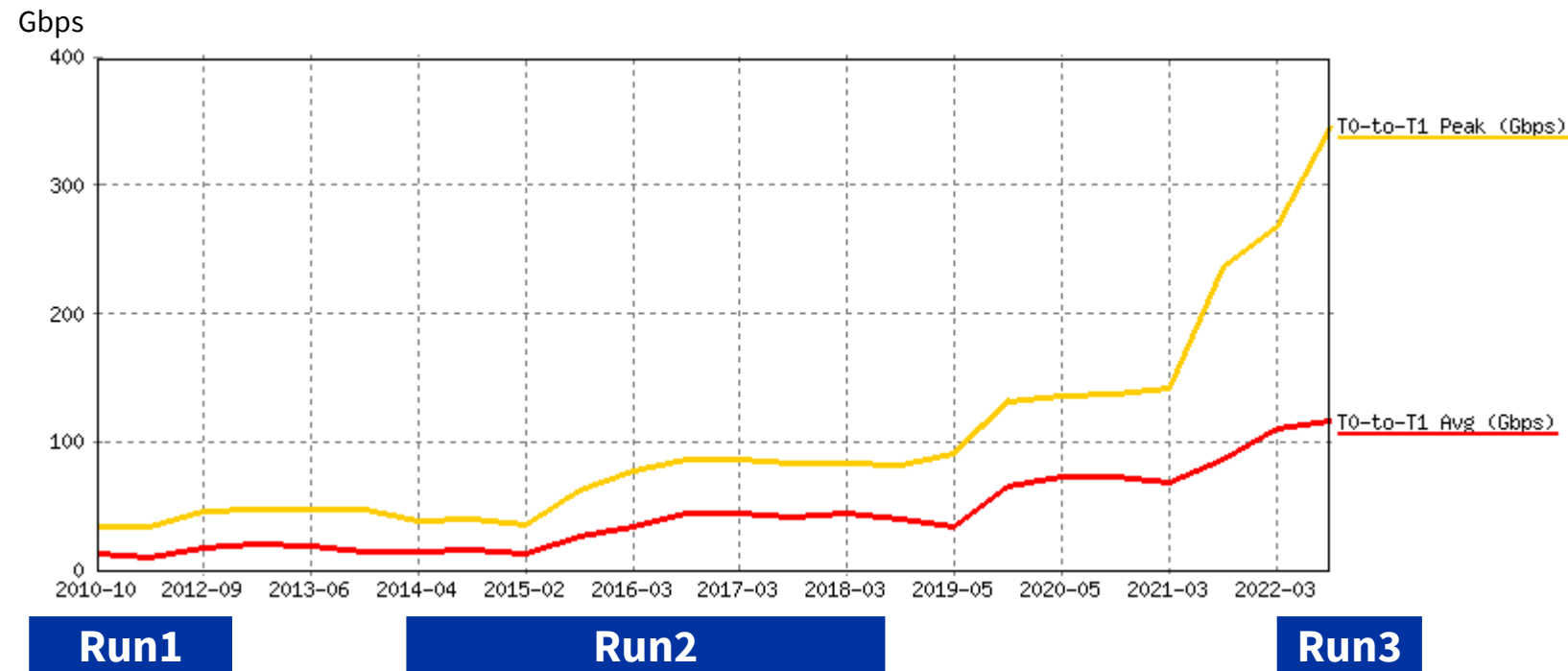
Networks are ready to support much larger volumes

Y-Axis: Gbps - Average LHCOPN bandwidth of previous 12 months

Ref: https://twiki.cern.ch/twiki/bin/view/LHCOPN/LhcopnStats

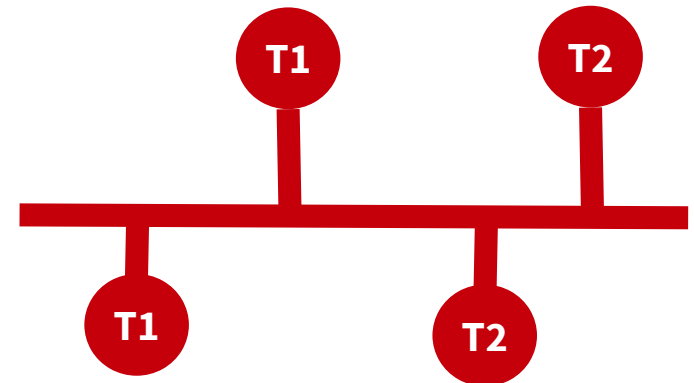# LHCONE

LHCONE L3VPN service

Private network connecting Tier1s and Tier2s

Secure:

- Dedicated to LHC data transfers
- Only declared IP prefixes can exchange traffic
- Can connect directly to Science-DMZ, bypass perimeter firewalls

Advanced routing:

- Multi domaina L3 VPN
- BGP communities for traffic engineering

Open to other HEP collaborations
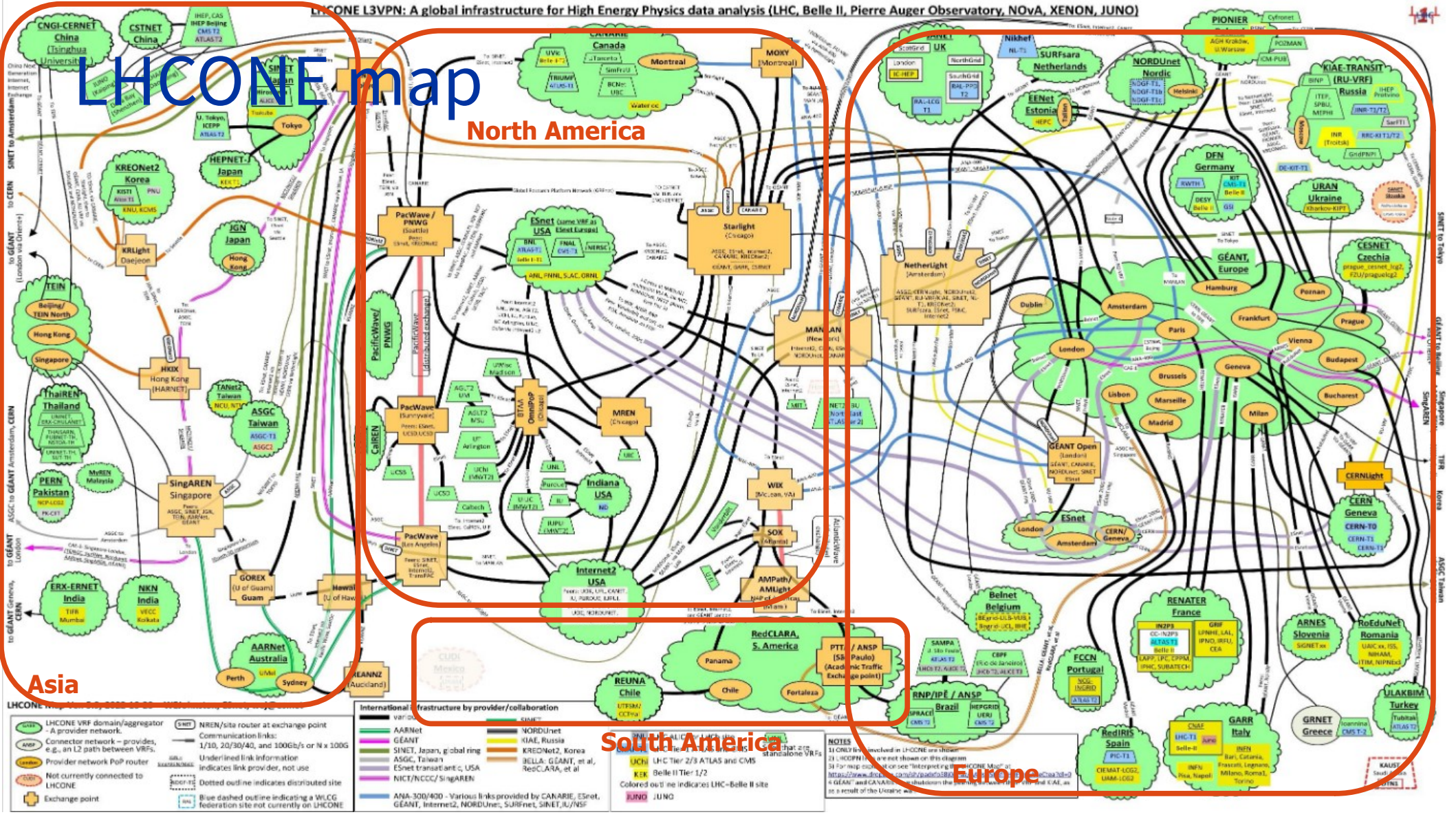
LHCONE L3VPN – latest news

News

- SURF LHCONE instance being implemented
- GARR is upgrading the Italian LHCONE backbone to 400Gbps
- DFN has now a 200Gbps connection to GEANT LHCONE
- CERN is going upgrading the GEANT LHCONE connection to 2x 400Gbps

Traffic statistics:

- Slovenia traffic has grown considerably, due to its HPC intensively used by WLCG
- LHCONE traffic has increased 30-40% in the last year in all regions
- More IPv6 traffic than IPv4 in all the LHCONE instances

LHCONE L3VPN: A global infrastructure for High Energy Physics data analysis (LHC, Belle II, Pierre Auger Observatory, NOvA, XENON, JUNO)

Asia-Pacific VRFs – Current Status

Internet2
US AS11537

ESnet
US AS293

CANARIE
CA AS6509

AARnet
AU AS7575

Transpac
US AS22388

MyREN
MY

ThaiREN
TH AS24475

SINET
JP AS2907

CERnet
CN AS23911

JGN
JP AS17934

CERNlight
CH AS20641

CSTnet
CN AS7497

ASGCNet
TW AS24167

NORDUnet
EU AS2603

IdREN
ID

TEIN
HK: AS24490
SG: AS24490

PERN
PK

KREONET
KR AS17579

NKN
IN AS55824

GEANT
EU AS20965

Asia-Pacific

Not Asia-Pacifc

LHCONE connection
LHCONE mutual transit
Connection in place, but no LHCONE
Missing connection

GEANT update

Developing its next generation network:

- **increasing dark-fibre footprint**

- reducing power consumption ~20%

- reducing cost ~40%

- 15+ years fibre leases

- partially disaggregated Optical Network

- offering wavelength to NRENs

- tender for new IP/MPLS routers starting now
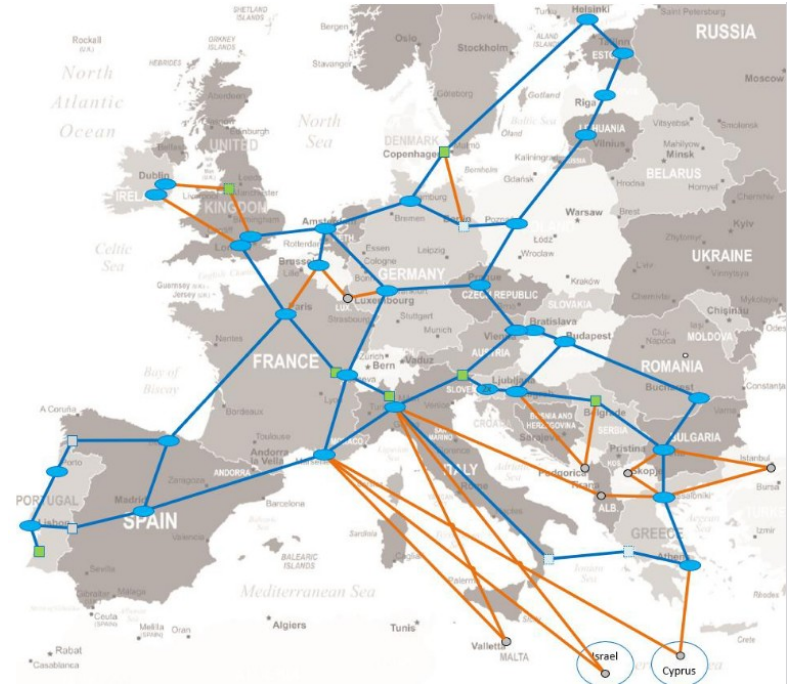
ESnet update

ESnet6 has been officially launched

Upgrades of EU links to 400Gbps is on-going

Trans-Atlantic capacity targets:
– 500G now
– 1.5T in Q3 2023
– 3.2T in 2027, well in advance of Run 4

US sites:
- BNL: now 300Gbps, soon 800Gbps
- FNAL: now 400Gbps, soon 800Gbps
- Tier2s will be upgraded to 400Gbps by 2027
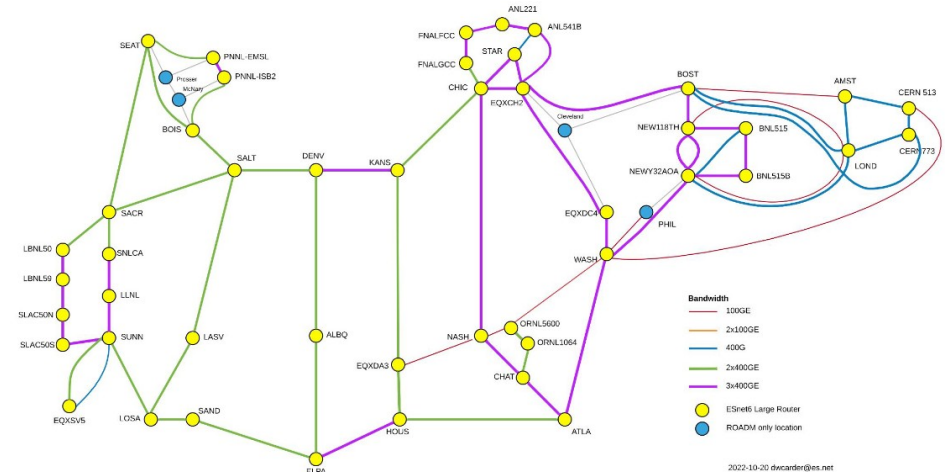
LHCONE monitoring

perfSONAR

- perfSONAR 5 beta is out and being tested. Still some stability issues due to the scale of the deployment and number of tests. v5 has been delayed to 2023
- perfSONAR 5 will use Elasticsearch and Grafana
- Infrastructure: the message bus is being phased out and data will be sent directly to Elasticsearch
- 100G mesh: data is now shown correctly , but results are not great. Work in progress
- Sites need to plan to update hardware as well as keeping the perfSONAR software updated. Especially needed for Data Challenge 2024 (DC24)
- pS-Dash: in development. Implemented AS traceroute: it hides the noise caused by load-balancing
- Total IN/OUT bandwidth: sites asked to provide URL with json of total in/out network counters. URLs stored in CRIC

LHCONE Looking Glass

Running looking-glass to analyses the routing tables of the VRFs

Implemented on a CERN router. Now peering with these VRFs:
- ASGC AS24167
- CANARIE AS6509
- CERNlight AS20641
- ESnet AS293
- KREOnet AS17579
- GEANT AS20965 (Geneva and Frankfurt routers)
- NORDUnet AS2603
- RU-VRF AS57484

The looking glass is accessible at http://lhcone-lg.cern.ch/

CRIC database

Database for WLCG network information

The LHCONE community has expressed the need for an authoritative database
to store network information, as:
- LHCONE and LHCOPN prefixes
- Autonomous System Numbers (ASN)
- Bandwidth
- LHCONE AUP ack
- perfSONAR references
- Monitoring links

CRIC

The CRIC (Computing Resources Information Catalogue) database has
    been chosen to store this information.

CRIC is the on-line database for WLCG Computing Resources:
http://wlcg-cric.cern.ch/
- Used by ATLAS and CMS, other experiments may follow
- not-WLCG sites can already be added

Database tables for network information
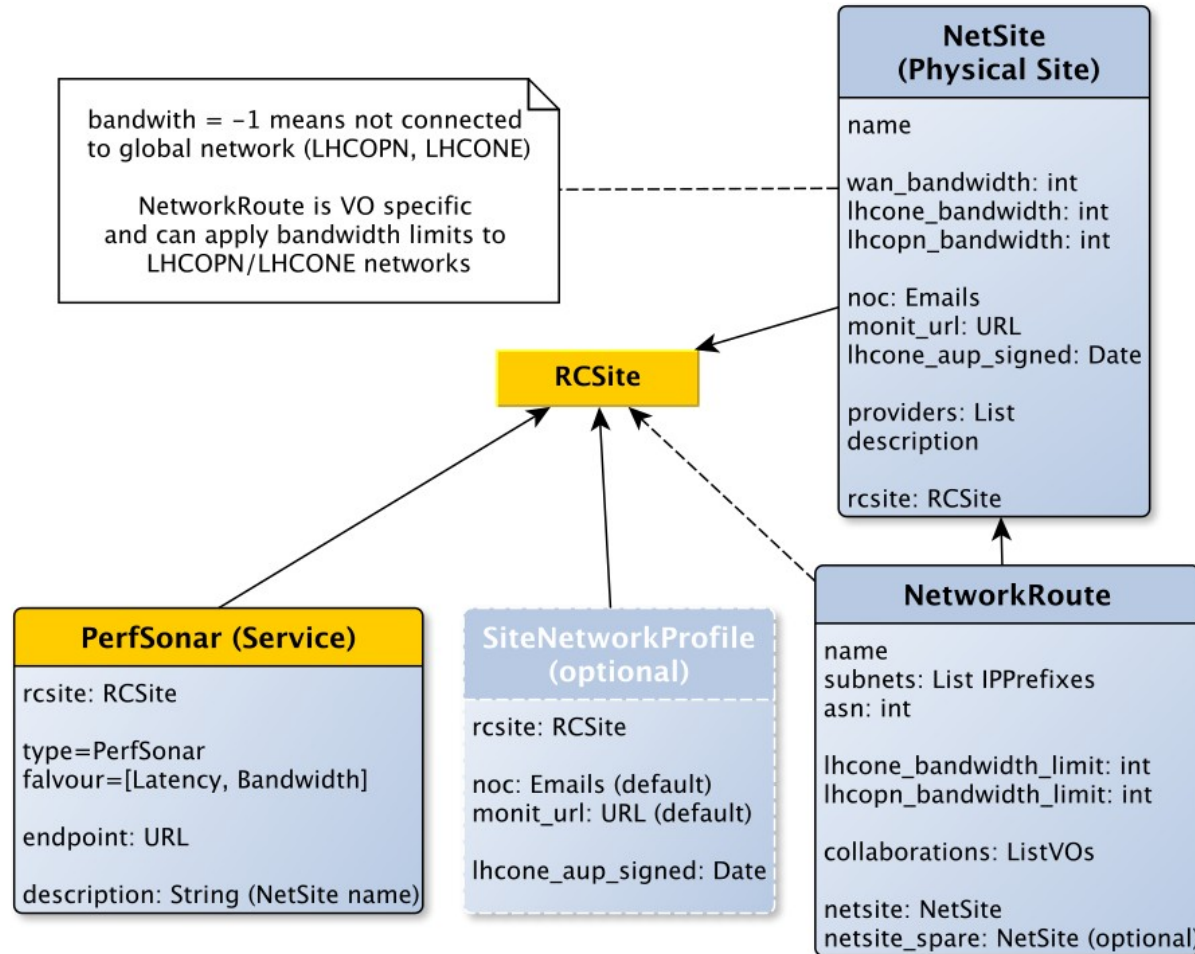
Access and Data

Easily accessible

- Netsite: https://wlcg-cric.cern.ch/core/netsite/list/ (login required)
- NetworkRoute: https://wlcg-cric.cern.ch/core/networkroute/list/ (login required)
- Json view: https://wlcg-cric.cern.ch/api/core/rcsite/query/?json (no login)

Data

- Database pre-loaded with data from LHCONE Twiki and LHCONE routing tables and verified by site administrators
- REN prefixes also added
- Sites can update their own data

NetSite table

https://wlcg-cric.cern.ch/core/netsite/list/

- NOC contact email
- URLs to monitoring and other info pages
- LHCONE AUP acknowledgment
- LHCONE and LHCOPN participation
- Network bandwidth: WAN, LHCOPN, LHCONE (-1 = not connected)
- LHCONE providers

~140 entries

RC Site	NetSite	NOC	monit URL	info URL	AUP	LHCONE active	LHCOPN active	WAN	LHCONE	LHCOPN	providers
AGLT2	US-AGLT2 Michigan State University	aglt2-noc@umich.edu			✔	✔	✖	100	100	-1	ESnet
AGLT2	US-AGLT2 University of Michigan	aglt2-noc@umich.edu			✔	✔	✖	80	80	-1	ESnet
ANLASC	US-ANL	noc@anl.gov			✔	✔	✖	0	100	-1	ESnet
ARNES	SL-ARNES-NREN				✔	✔	✖	0	9	-1	ARNES
ARNES	SL-IJS-Ljubljana				✔	✔	✖	0	9	-1	ARNES
ARNES	SL-IZUM-Maribor				✔	✔	✖	0	9	-1	ARNES
Australia-ATLAS	AU-Australia-ATLAS				✔	✔	✖	20	10	-1	AARNET
BEgrid-ULB-VUB	BE-ULB-VUB				✖	✖	✖	20	-1	-1	

NetworkRoute table

https://wlcg-cric.cern.ch/core/networkroute/list/

One entry per set of prefixes sharing common routing policies

- set of v4 and v6 prefixes
- "More specific" flag (in case of disaggregated prefixes)
- Autonomous System number (ASN)
- collaborations using these prefixes
- LHCOPN and LHCONE bandwidth specific to the prefixes of the record
- monitoring URL specifics to the prefixes of the record

~150 entries

RC Site	NetworkRoute	NetSite	ASN	monit URL	MS	Subnets	LHCONE limit	LHCOPN limit	collaborations
AGLT2	✏️⊞ AGLT2_LHCONE_RT	US-AGLT2 Michigan State University	229		✗	2001:48a8:68f7::/48, 2001:48a8:68f7::/50, 2001:48a8:68f7:4000::/50, 2001:48a8:68f7:8000::/50, 2001:48a8:68f7:c000::/50	100	-1	US-ATLAS, WLCG
AGLT2	✏️⊞ AGLT2_MSU	US-AGLT2 Michigan State University	229		✗	192.41.236.0/23, 192.41.238.0/28	100	-1	US-ATLAS, WLCG
AGLT2	✏️⊞ AGLT2_MSU IPv6	US-AGLT2 Michigan State University	237		✗	2001:48a8:68f7:8001::/64	100	-1	US-ATLAS, WLCG

CRIC vs LHCONE TWIKI

CRIC is now the authoritative database for the list of LHCONE connected sites and LHCONE prefixes

List on LHCONE Twiki page has been removed

LHCOPN data has also been added

RS-LHCONE route-set

RS-LHCONE: RIPE database route-set object with the list of all the LHCONE
    prefixes declared in CRIC

The route-set object is stored in the RIPE whois database:
    https://apps.db.ripe.net/db-web-ui/query?searchtext=rs-lhcone
    `>whois -h whois.ripe.net RS-LHCONE`

A script that generate the route-set runs every day:
    - Manual update of the route-set
    - Upon updates, emails can be sent to interested NOCs using the "notify:"
    attribute. Tell Edoardo Martelli if interested

RS-LHCONE route-set

```
whois -h whois.ripe.net rs-lhcone

% Information related to 'RS-LHCONE'

route-set:      RS-LHCONE
descr:          List of prefixes available in LHCONE
tech-c:         LHC1
admin-c:        LHC1
mnt-by:         LHCONE-MNT
mnt-by:         CERN-MNT
created:        2020-07-07T12:35:15Z
last-modified:  2022-02-22T13:27:38Z
source:         RIPE
members:        109.105.124.0/22  # AS39590 - NDGF-T1-LHCOPNE
members:        117.103.96.0/20   # AS24167 - Taiwan-LCG2-LHCOPNE
members:        128.104.227.0/24  # AS59 - GLOW-LHCONE
members:        128.111.120.96/28  # AS131 - US-Univ-California-Santa-Barbara-UCSB-LHCONE
members:        128.142.0.0/16  # AS513 - CERN-PROD-LHCOPNE
members:        128.211.128.0/19  # AS397511 - Purdue-LHCONE
members:        128.211.160.0/22  # AS397511 - Purdue-LHCONE
members:        128.227.10.0/24  # AS6536 - UFlorida-HPC-LHCONE
...
mp-members:     2a05:81c5:302:224::/64  # AS2875 - JINR-LCG2-LHCONE
mp-members:     2a05:81c5:302:37::/64  # AS2875 - JINR-LCG2-LHCONE
mp-members:     2a07:8504:120:e060::/64  # AS1104 - NLT1-NIKHEF-LHCOPNE
mp-members:     2a07:8504:120:e068::/64  # AS1104 - NLT1-NIKHEF-LHCOPNE
mp-members:     2a07:a6c0:b:192::/64  # AS52146 - RU-SARFTI-LHCONE
mp-members:     2a0c:5bc0:c8:2::/64  # AS786 - UKI-LT2-IC-HEP-LHCONE
mp-members:     2a0e:e140::/64  # AS58255 - Ru-Troitsk-INR-LCG2-LHCONE
```

Use of RS-LHCONE to build BGP filters

All prefixes in the LHCONE routing table matches an entry in CRIC
    - several checks in January and February 2022
    - excluded prefixes have been fixed with site admins

BGP filters generated from RS-LHCONE can be safely applied
    - is anyone planning to do it?

Resources:

Documentation and example scripts:

https://twiki.cern.ch/twiki/bin/view/LHCONE/LhcOneVRF#Connected_Sites

Data challenges

WLCG Data Challenges

WLCG Data Challenge 2021

- Achieved expectations (10% of HL-LHC)

- Network not saturated, but somehow stressed at exchange points

Next data challenge (DC24)

- Originally planned for 2023,now delayed to Spring 2024

- Target of 30% of HL-LHC requirements 30% is 3x increase and sites may not have
    enough hardware. Target may is being reviewd

- Network providers have already planned upgrades. Network capacity should be
    enough for DC24

- The community needs to plan what is necessary to be tested

DC24 Planning

The last LHCONE meeting was followed by a dedicated meeting for the preparation of the next WLCG Data Challenge in 2024

Important points discussed:
- This period before the challenge should be used to pre-test individual components with mini-challenges and milestones
- The duration of DC24 should be extended to allow to understand possible problems and also to allow to test different scenarios
- New network functionalities like packet marking and SDN projects should be part of the challenge
- Sites should not buy hardware just to meet DC24 requirements (the hardware may be obsolete by the time of Run4). Goal of 30% may be reviewed

Notes from the meeting here
The meeting was also discussed at the WLCG workshop. Shawn's presentation is here

Other HEP Collaborations and Big Sciences

Other collaborations and sciences

During last meeting:

- Updates from Juno and BelleII, already members of LHCONE

- Invited talks from ITER and SKA, big science projects which in the future may compete with WLCG on network utilization

- DUNE was also discussed: its sites are already members of LHCONE and it may formalize the use of LHCONE by signing the AUP

There's a growing need for Bic Science collaborations to coordinate their requirements to allow an organic grow of the R&E networks

MultiONE ideas will be revamped

multiONE

LHCONE already very large, it could become risky to include other large science projects

Better to implement multiple VPNs, one for each collaboration:
- Each site joins only the VPNs it is collaborating with, to reduce the exposure of their data-centre
- Each Collaboration funds its own VPN

But it's difficult to separate the traffic for sites member of multiple collaborations. Work in Progress
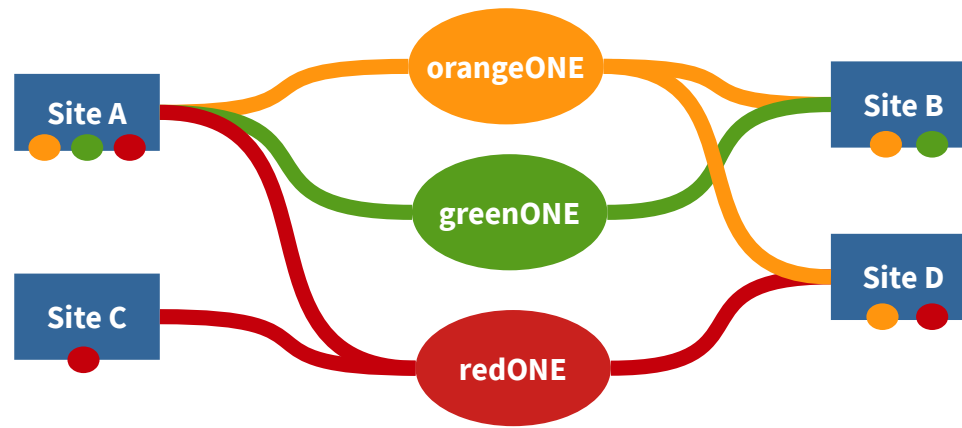
WLCG guidelines

Message from Simone Campana:

In the next 10 years WLCG will be faced with two major network challenges:
- dealing with the HL-LHC data volumes and complexity
- the cohabitation with other experiments and sciences on the same infrastructure

The network community can play a leading role:
- modernize the network services, progressing with the ongoing R&D activities and
    bringing early prototypes in production
- engage with other experiments and sciences to drive the evolution of R&E
    networks

Network R&D

NOTED

NOTED is a framework that can detect large FTS data transfers and trigger network optimization actions to speed up the execution of the transfers

Already tested with production transfers:
- CERN-PIC with LHCOPN-LHCONE load balancing
- CERN-TRIUMF and KIT-TRIUMF with the activation of dynamic circuits
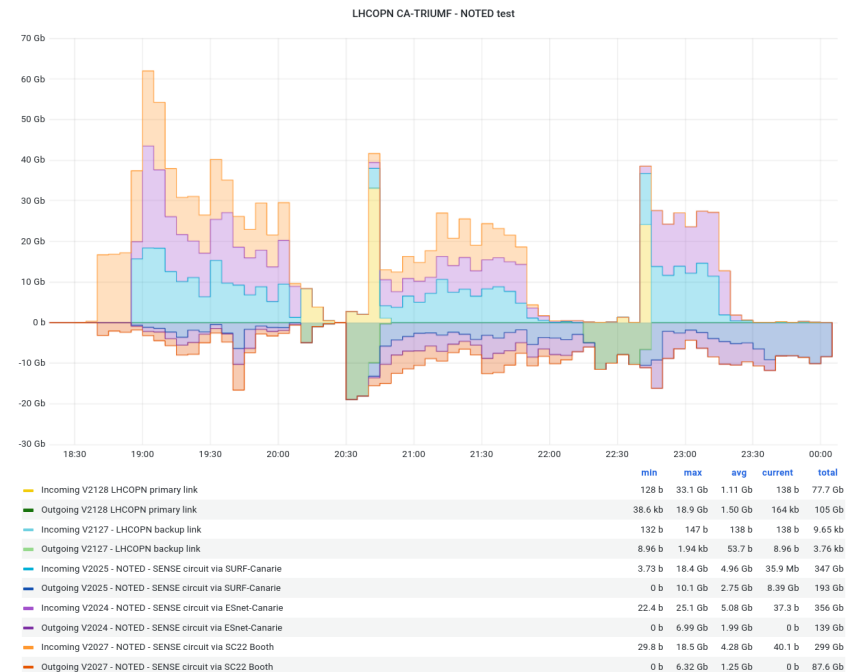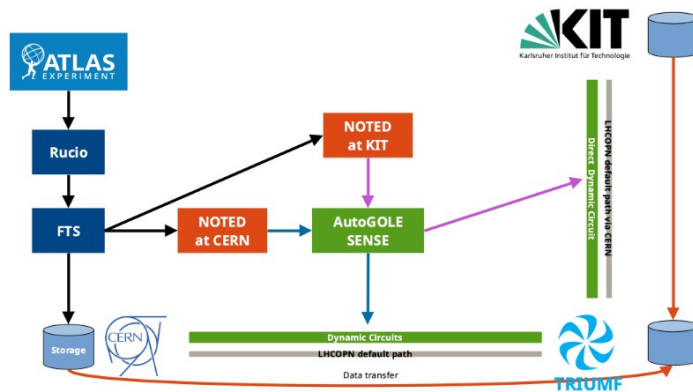
NOTED 2.0
- python code is now available on PiP for download.
- Evaluating possible integration with FTS to tu the FTS Optimizer when more bandwidth i provisioned.
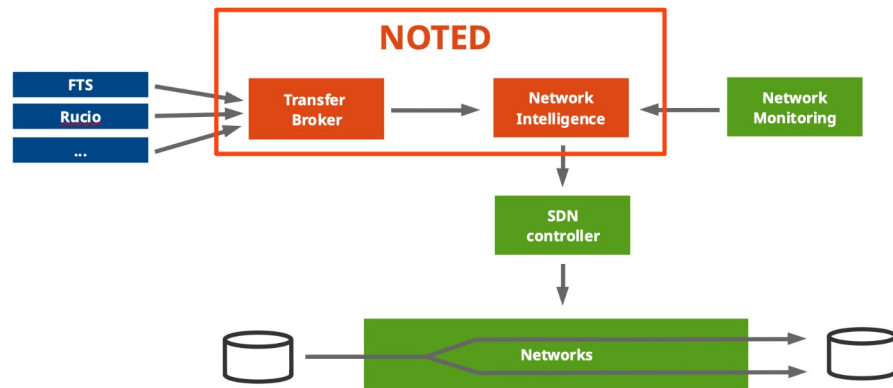
NOTED demo at SC22

A demo was run during SC22:
- two NOTED instances, one a CERN and one at de-KIT
- provisioning SENSE dynamic circuits from CA-TRIUMF,
    one going via the SC22 Starlight Booth
- successfully detected transfers

LHCOPN CA-TRIUMF - NOTED test

	min	max	avg	current	total
Incoming V2128 LHCOPN primary link	128 b	33.1 Gb	1.11 Gb	138 b	77.7 Gb
Outgoing V2128 LHCOPN primary link	38.6 kb	18.9 Gb	1.50 Gb	164 kb	105 Gb
Incoming V2127 - LHCOPN backup link	132 b	147 b	138 b	138 b	9.65 kb
Outgoing V2127 - LHCOPN backup link	8.96 b	1.94 kb	53.7 b	8.96 b	3.76 b
Incoming V2025 - NOTED - SENSE circuit via SURF-Canarie	3.73 b	18.4 Gb	4.96 Gb	35.9 Mb	347 Gb
Outgoing V2025 - NOTED - SENSE circuit via SURF-Canarie	0 b	10.1 Gb	2.75 Gb	8.39 Gb	193 Gb
Incoming V2024 - NOTED - SENSE circuit via ESnet-Canarie	22.4 b	25.1 Gb	5.08 Gb	37.3 Gb	356 Gb
Outgoing V2024 - NOTED - SENSE circuit via ESnet-Canarie	0 b	6.99 Gb	1.99 Gb	0 b	139 Gb
Incoming V2027 - NOTED - SENSE circuit via SC22 Booth	29.8 b	18.5 Gb	4.28 Gb	40.1 b	299 Gb
Outgoing V2027 - NOTED - SENSE circuit via SC22 Booth	0 b	6.32 Gb	1.25 Gb	0 b	87.6 Gb

48

# Research Network Technology WG - update

The RNTWG has made significant progress on network traffic visibility through the work on IPv6 flowlabel tagging and Firefly flow marking

Flowd: flow and packet marking service developed in Python
- supports plugins to get connections which have to be marked
- backends are used to make the marking and other tasks
  - eBPF backend for flowlabel tagging
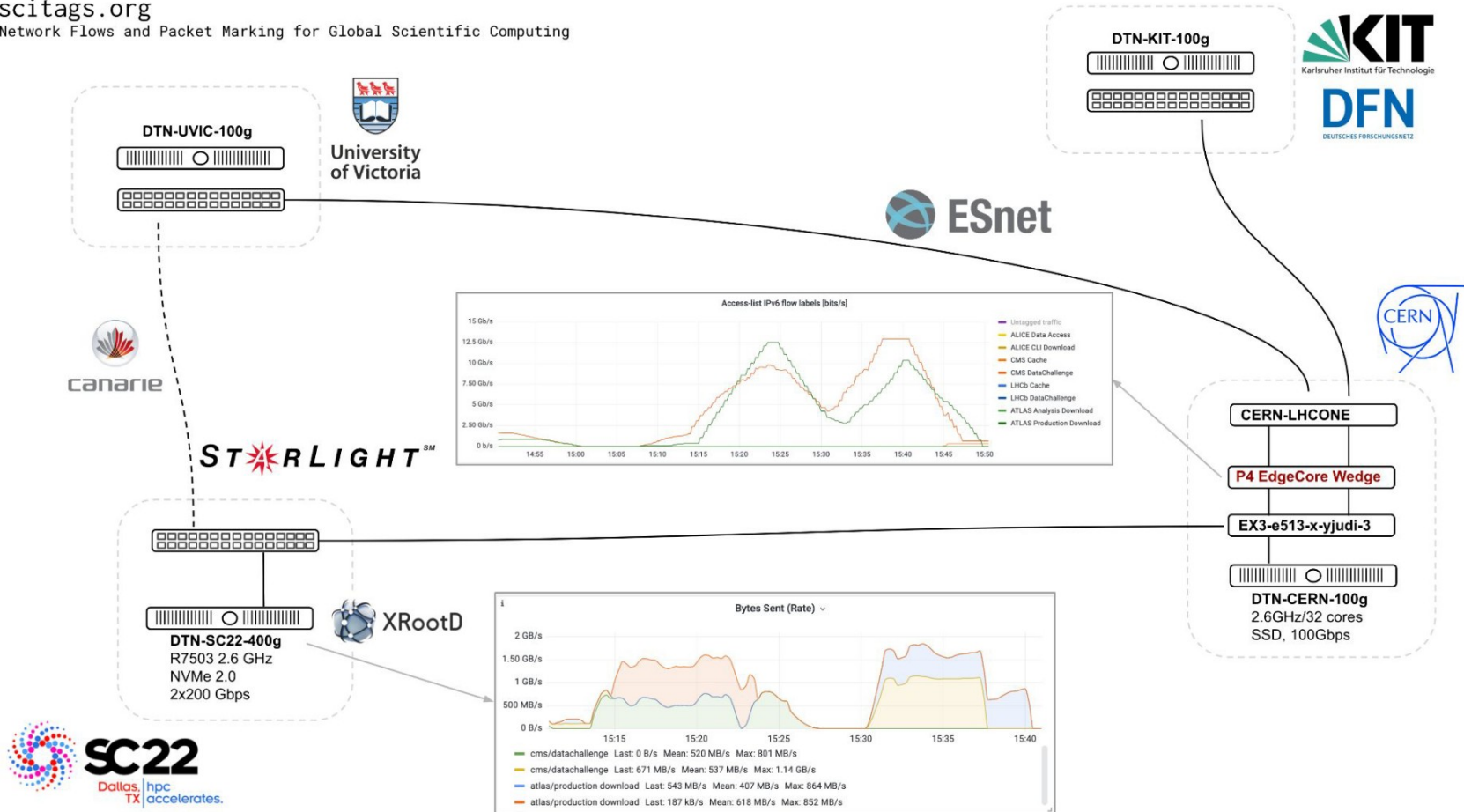 - Prometeus backend to expose marked connections

scitags.org

How scitags works
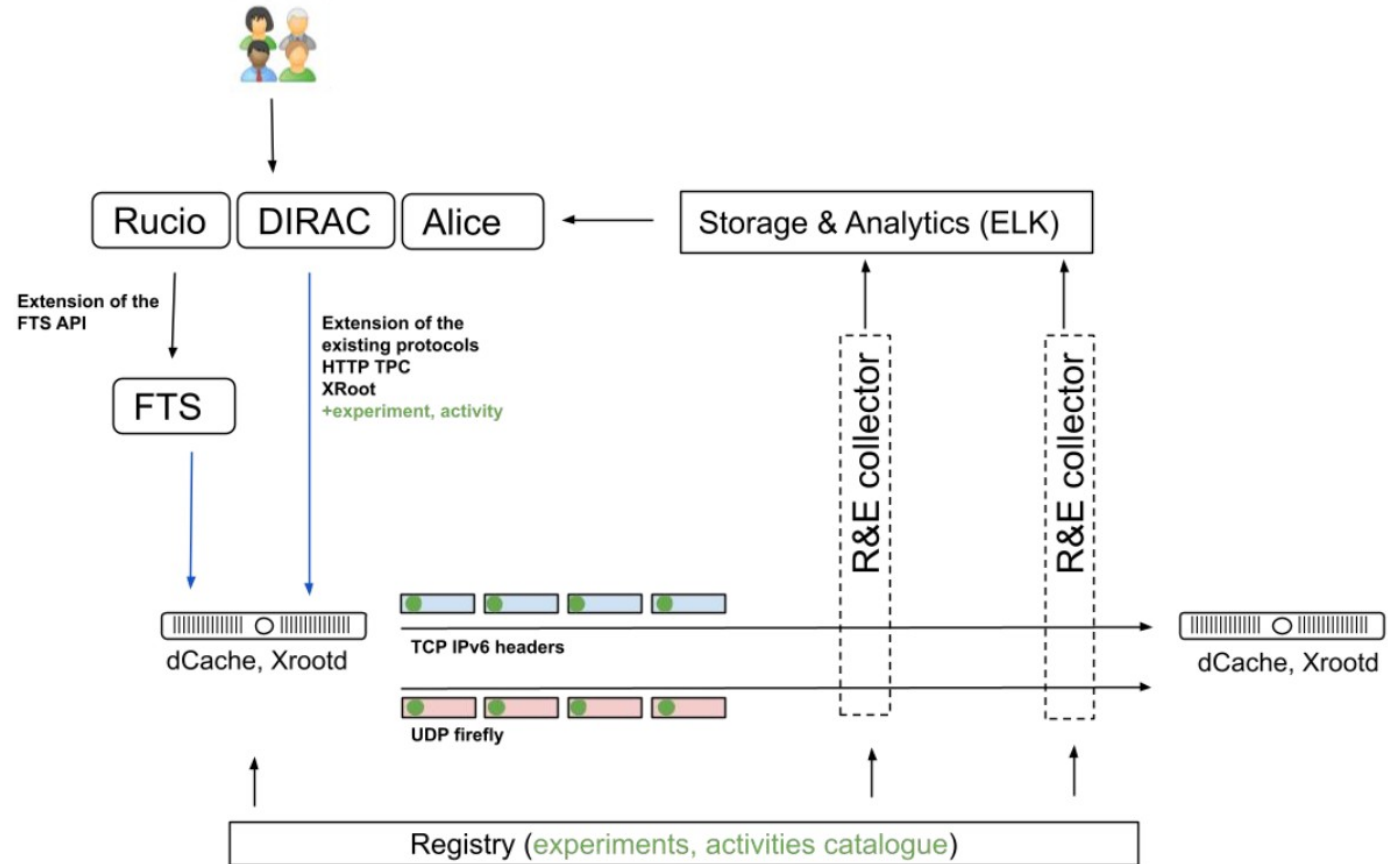
Packet tagging demo at SC22

AutoGOLE and SENSE

AutoGOLE: Infrastructure which provides "end-to-end" network services in a fully automated manner

Open-source software framework based on:

- Network Service Interface (NSI): multidomain network provisioning

- SENSE: end-system provisioning and realtime integration with network services

Persistent Infrastructure, somewhere in between production and a testbed

AutoGOLE, NSI and SENSE work together to provide the mechanisms for complete end-to-end services that include network and attached End Systems DTNs

Circuit provisioning functionality used by NOTED and Scitags demo for SC22

Using SENSE to move CMS data in Rucio

Project led by UCSD and Caltech

The increased requirements of the HL-LHC requires to use
any resource in the most efficient way, including networks

Objectives of the project:

#1 Make Rucio capable to schedule transfers on the network and prioritize them

#2 Predetermined transfer speed and quality of service (time to completion)

Demonstrated:

- SENSE can build VPNs between pairs of XrootD servers in charge of FTS transfers
requested by Rucio

- QoS can be provisioned in the network to prioritize the traffic in the VPN

Wish to participate to DC24
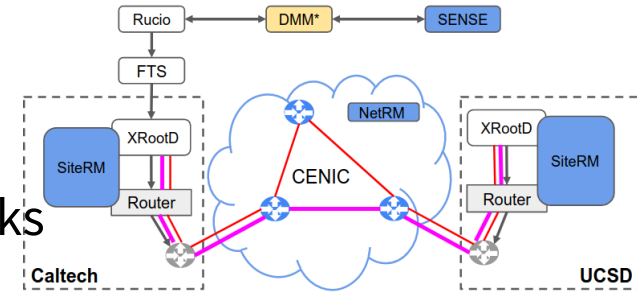
CNAF-CERN DCI

Pilot for a Tbps Data-Centre interconnections between INFN-CNAF and CERN

- to meet HL-LHC requirements at an affordable cost

- made using transmission devices at CNAF and CERN

  - no need to use expensive routers

- using optical channels on GARR and GEANT dark fibres

  - pilot also for their new Spectrum Sharing service

  - distance (~1000km) is also a challenge

- GARR will procure and manage the
   Infinera devices at both
   ends of the connection

- Equipment installed and connected on 16th/11.
   -Reached 400Gbps on single wavelength Geneva-Bologna
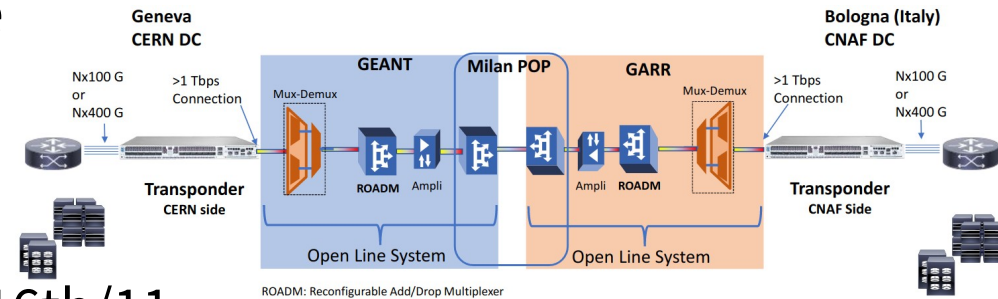
Conclusions

Summary

- Run3 is on-going, experiments are collecting new data
- LHCOPN and LHCONE grows as traffic keeps increasing
- CRIC now used also to store network information
- Preparation of Data Challenge 2024 has started
- WLCG needs to prepare to share the network with upcoming sciences
- Network R&D essentials to be ready for Run4 data deluge

Next LHCOPN-LHCONE meeting

Options: Amsterdam and Catania. Being voted.
April 2023
In person meeting

Agenda will be published here
https://indico.cern.ch/e/LHCOPNE50

Questions?

edoardo.martelli@cern.ch